Date: February 9, 2006

To: Teresa Michelsen, Avocet Consulting

From: Mike Anderson, DEQ

cc: Jennifer Peterson and Mike Poulsen, DEQ-NWR

Re: Follow-up to our 1/30/06 meeting

One of the tasks that I wanted to accomplish after our 1/30 meeting was to look at the macro codes of the DEQ and LWG spreadsheets to see if I could determine why the DEQ spreadsheet consistently yields higher values of % false positives (%FP) than does the LWG spreadsheet. Instead of just looking at the specific place in the code where this parameter is calculated, I looked at all of the macros to see if there were program differences from which these %FP differences propagate. The bullets below summarize my observations on similarities or differences between the DEQ and LWG approach for each step in the Floating Percentile calculations.

Make Data Table:

- The DEQ and LWG spreadsheets generate the exact same data table.

Create Distributions:

- DEQ uses all the available data – hit, no-hit, and indeterminate – to create a single distribution for all of the detected concentrations for each chemical.
- LWG creates two concentration distributions for each chemical; one from the hit data and one from the no-hit data. The indeterminate data are not used in either distribution.
- Because of these differences the two distribution tables could not be compared directly. Basic comparisons, however, like checking to see if each spreadsheet correctly identified the same chemicals as hits and no-hits showed that there were no differences at that level.

Calculate Percentiles:

- DEQ calculates one set of percentiles for each chemical based on the single distribution created above.
- LWG creates two sets of percentiles for each chemical, one based on the hit distribution and the other on the no-hit distribution. The indeterminate data are not used in the LWG spreadsheet.
- DEQ uses a built-in Excel function to calculate the percentiles. LWG has written a macro to estimate the percentiles. Test calculations comparing these two methods showed that they do not generate exactly the same values for the percentiles, but the differences are very slight and are not expected to have any noticeable affect on the results.

- As would be expected, data sets with more overlap, that is, where the ranges of hit and no-hit are somewhat similar, yield LWG percentiles that are closer to each other and closer to the DEQ percentiles. Data sets with little overlap show larger differences. Since the DEQ percentiles include all of the data its percentiles fall between the no-hit and hit distributions. The table below shows how these distributions vary for five different compounds in the CHG-L2 data set.

	Cadmium	**Dieldrin**	**Diesel Range Hydrocarbons**	**Di-n-butyl phthalate**	**Lead**
	10th Percentile				
No-Hit (LWG)	0.11	0.12	29.90	4.80	7.91
Combined (DEQ)	0.11	0.12	35.10	4.90	8.33
Hit (LWG)	0.15	1.38	335.00	88.20	13.60
	50th Percentile				
No-Hit (LWG)	0.24	0.32	110.00	10.00	17.20
Combined (DEQ)	0.25	0.40	134.00	11.00	18.60
Hit (LWG)	0.34	3.53	615.00	160.00	34.20
	90th Percentile				
No-Hit (LWG)	0.60	4.01	363.00	63.00	54.80
Combined (DEQ)	0.71	6.18	1066.00	97.50	68.50
Hit (LWG)	0.87	250.36	5810.00	1260.00	427.30

- If the LWG method sometimes uses the no-hit distribution and sometimes uses the hit-distribution, the LWG distributions may be lower or higher than the corresponding DEQ values and significant differences could result.

Error Calculations:

- Because the number of false positives (#FP) and percent false negatives (%FN) and other parameters are based on the results in from the percentile tables, the differences in the error calculations parallel those in the percentile calculations. DEQ has a single set, LWG has two different sets, and the differences can vary considerably. The resulting tables cannot be compared directly.
- The DEQ results range from 0% to 100% or vice versa in the appropriate columns, such as %FN and %FP. The LWG results, however, often do not reach these expected extreme values but end up several percent off in some data sets. This is another difference in the DEQ and LWG results, but the effect of this difference may be small.
- DEQ results have significantly higher #FP than the LWG results. Examination of the two spreadsheets did not turn up an obvious difference in the way this value is tabulated. When the LWG results are placed in the DEQ spreadsheet and the performance measures are calculated, all of the results agree with those obtained in the LWG spreadsheet. This shows that these parameters are calculated the same way in both spreadsheets. It also shows that the #FP is a sensitive parameter and that one or two values that "peak" in the DEQ data set and not in the LWG dataset or vice versa could lead to significantly different results.

Starting Criteria:

- Once again, the use of one versus two distributions can lead to varying degrees of difference. Since the LWG method has two distributions from which to choose, it selects two potential sets of data and then uses the one that has fewer #FP. In some cases the hit data are selected all of the time (CHP-L1), in some cases the no-hit data are selected all of the time (HYG-L2), and in some cases both are selected at various times during the calculation (HYP-L2).

Floating Percentiles:

- The performance measures like %FN, sensitivity, reliability, etc. are pulled from the percentile tables at the time that the starting criteria are selected. However, they do not appear to be recalculated on the basis of the final set of values generated by the FPM calculations. (This may be done manually for the LWG results since those values are adjusted manually at the end, but the DEQ spreadsheet does not recalculate the performance measures). If the calculation is carried out correctly there should be no effect on any results that are based on %FN since that parameter is held constant within a given test. However, after the %FN is selected the number of false positives is reduced. Therefore, the %FP should decrease from their initial value. Performance measures calculated from the final DEQ data sets were much closer to those reported by LWG. For example, in CHG-L1 the DEQ results for %FP ranged from 69% down to 40% for the five calculations. The LWG results ranged from 33% down to 20%. The recalculated performance parameters for DEQ had %FP values ranging from 43% to 25%, which are much closer to the LWG values.
- The DEQ macros are written so that %FN does not exceed any specified value. Examination of the LWG data shows that some of their results are for values that exceed the default 5%, 10%, *etc*. values used in the DEQ spreadsheet. For example, in the CHG-L1 calculations the LWG results are for the following percentages: 6, 12, 15, 21, and 26. The DEQ results are for 3, 9, 12, 18, and 24%. Different starting values for %FN should yield different results for the screening concentrations.

Although this is not a rigorous comparison of the two spreadsheets I believe that it is thorough enough to demonstrate that the differences between the DEQ %FP and the LWG %FP is not due to a basic error in programming the definition into the spreadsheets. The differences are real and appear to be due primarily to (1) the different distributions used for establishing the tables of percentiles, (2) the performance measures being based on the initial criteria and not on the final values (at least in the case of the DEQ spreadsheet), and (3) the five sets of calculations being generated from differing target values of %FN.